

Machine Learning 1

Lecture 1.2 - What is Machine Learning?

Erik Bekkers

(Bishop 1.0 and 1.1)



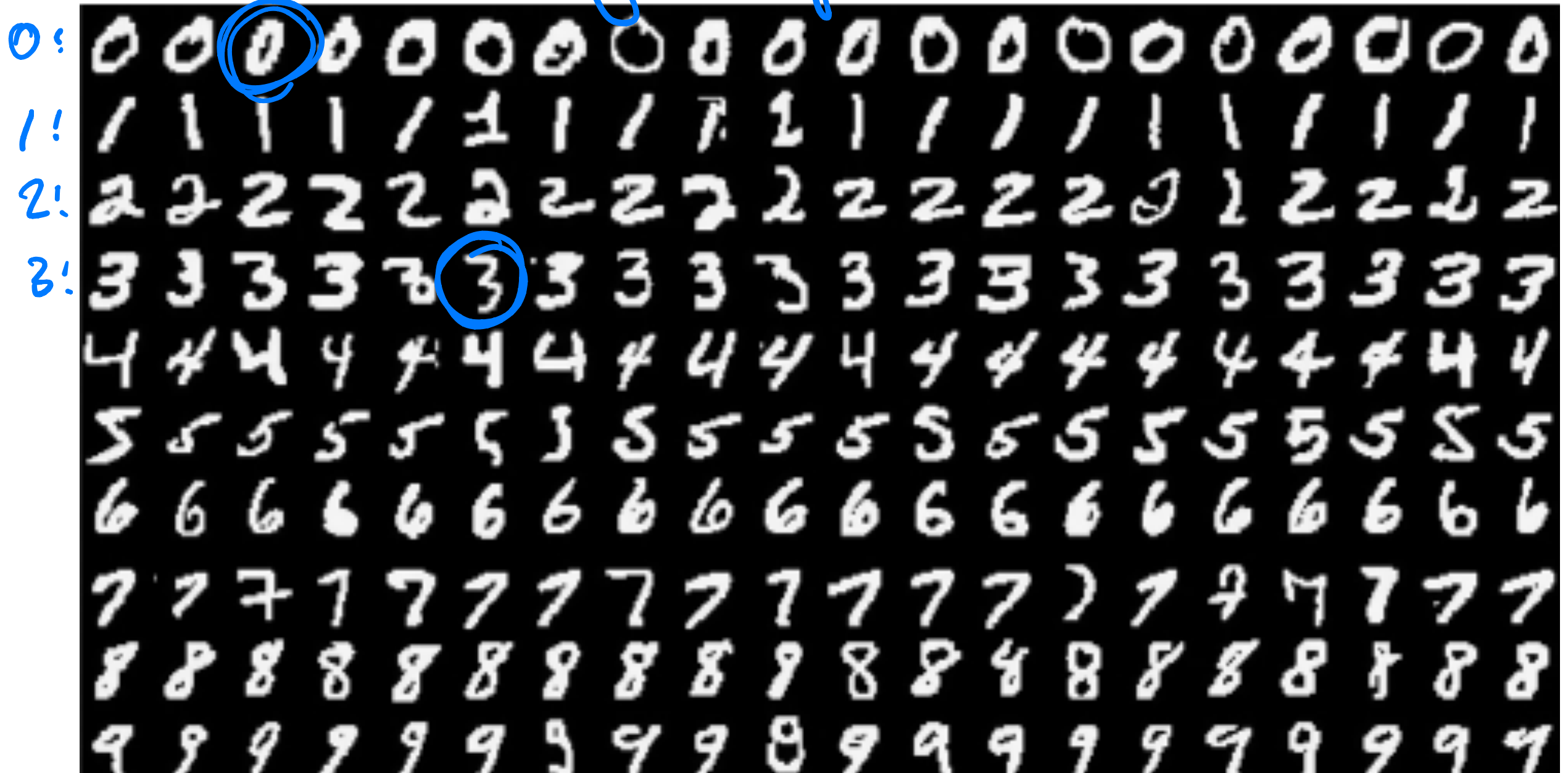
What is machine learning?

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .”

- Tom M. Mitchell

E: Experience

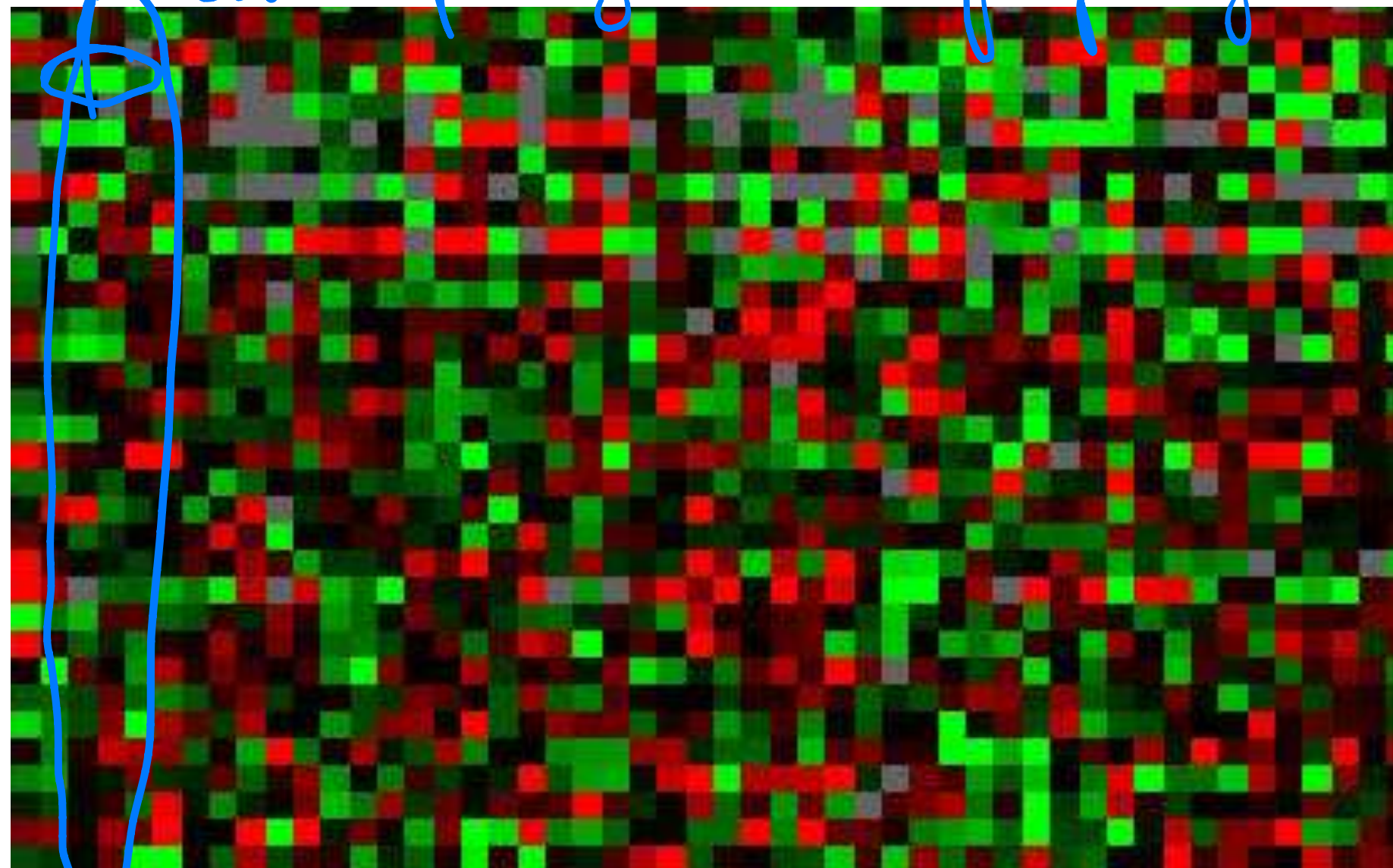
Handwritten digits of size 28×28



MNIST dataset

E: Experience

heat map for activity of a gene



ESTs
SIDW486740
SMALLNUC
ESTs
SIDW366311
SIDW357197
SID52979
ESTs
SID43609
SIDW416621
ERLUMEN
TUPLE1TUP1
SIDW428642
SID381079
SIDW298052
SIDW417270
SIDW362471
ESTsChr.15
SIDW321925
SID380265
SIDW308182
SID381508
SID377133
SIDW365099
ESTsChr.10
SIDW325120
SID360097
SID375990
SIDW128368
SID301902
SID31984
SID42354

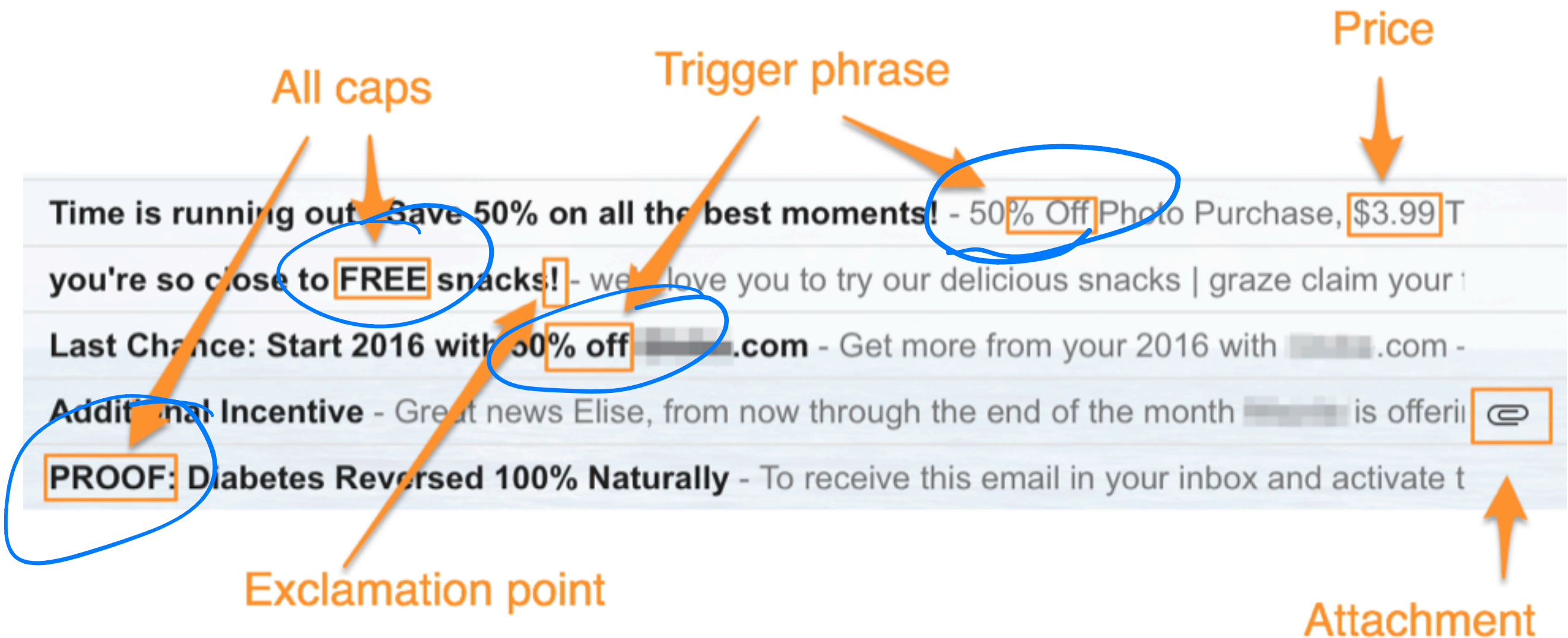
gene

BREAST
RENAL
MELANOMA
MELANOMA
MCF7D-repro
COLON
COLON
K562B-repro
COLON
NSCLC
LEUKEMIA
RENAL
MELANOMA
BREAST
CNS
CNS
RENAL
MCF7A-repro
NSCLC
K562A-repro
COLON
CNS
NSCLC
NSCLC
LEUKEMIA
CNS
OVARIAN
BREAST
LEUKEMIA
MELANOMA
MELANOMA
OVARIAN
OVARIAN
NSCLC
RENAL
BREAST
MELANOMA
OVARIAN
OVARIAN
NSCLC
RENAL
BREAST
MELANOMA
LEUKEMIA
COLON
BREAST
LEUKEMIA
COLON
CNS
MELANOMA
NSCLC
PROSTATE
NSCLC
RENAL
RENAL
NSCLC
RENAL
LEUKEMIA
OVARIAN
PROSTATE
COLON
BREAST
RENAL
UNKNOWN

Tumor

Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

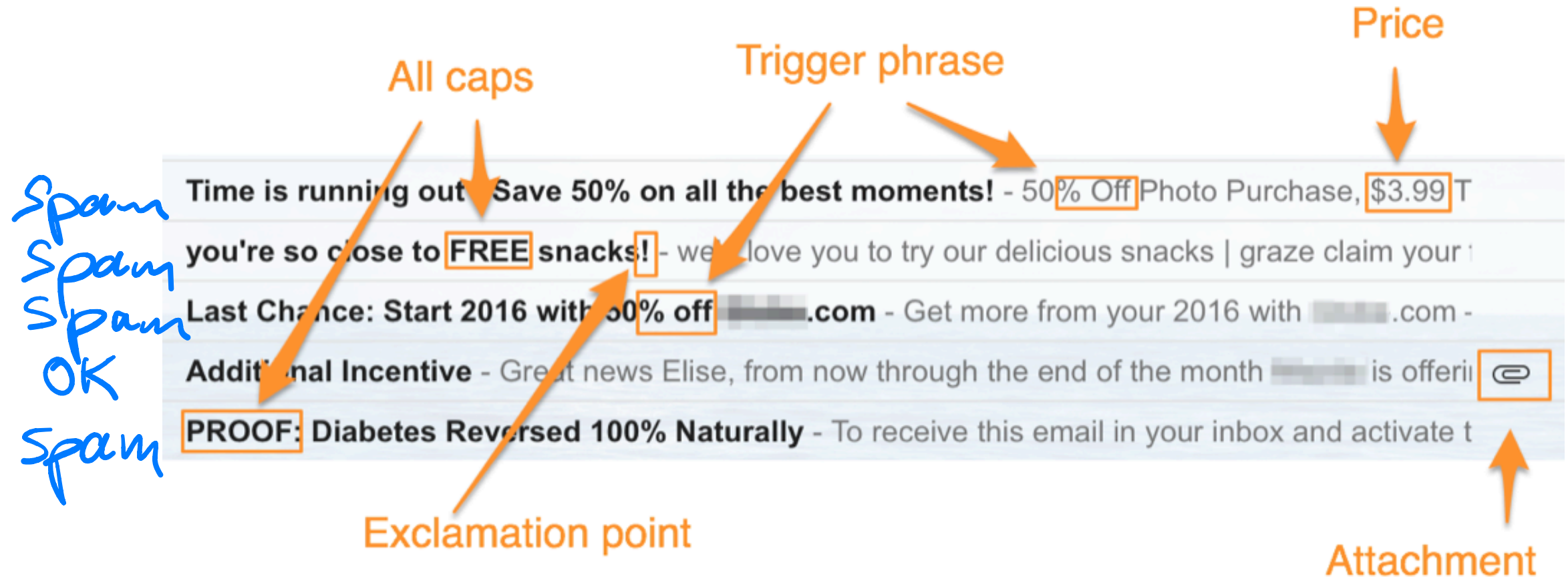
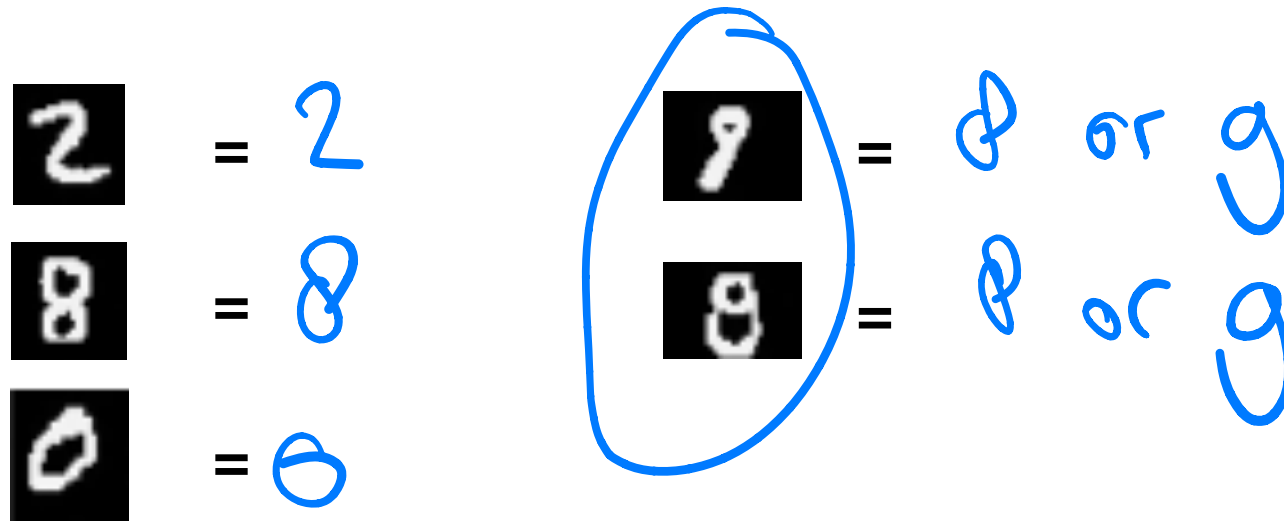
E: Experience



Examples of spam emails. [source: Yesware]

T: Class of tasks

Classification:



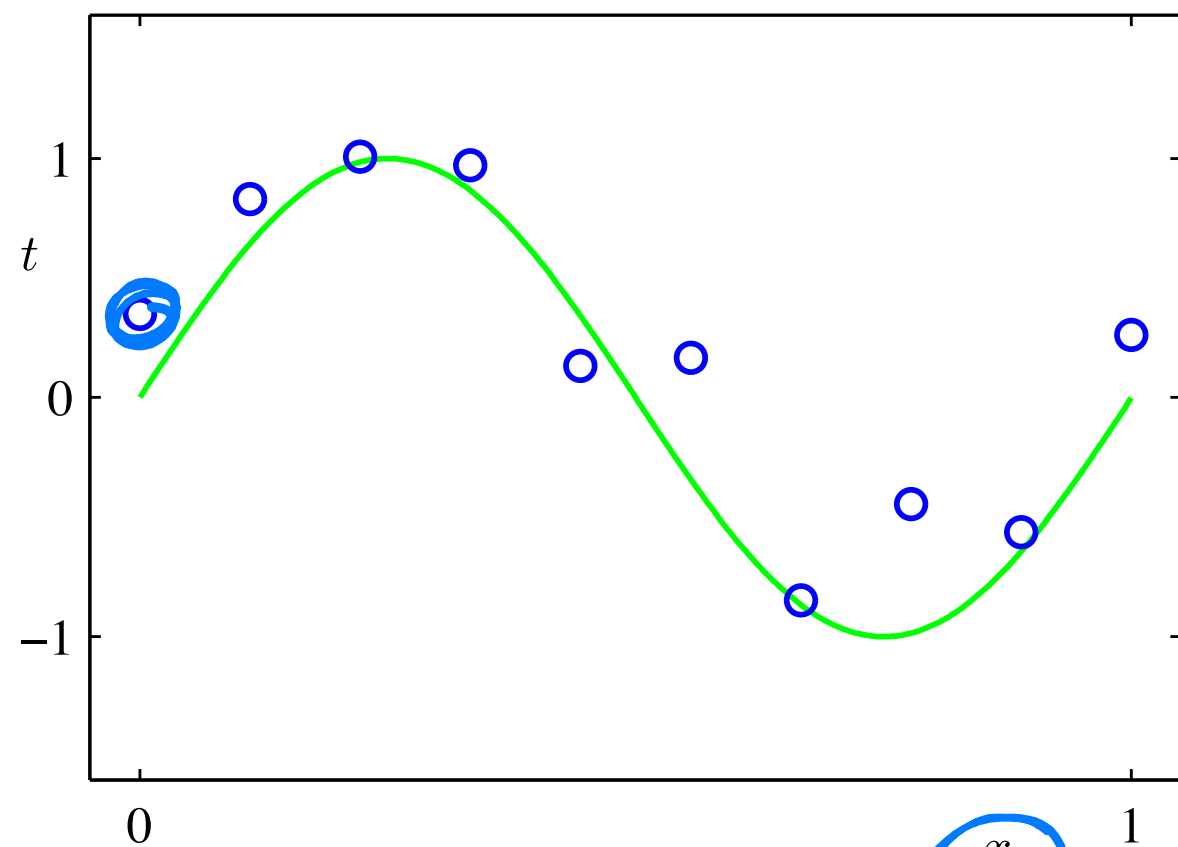
T: Class of tasks

Regression

input : x

target : $t = \sin(2\pi x) + \varepsilon$

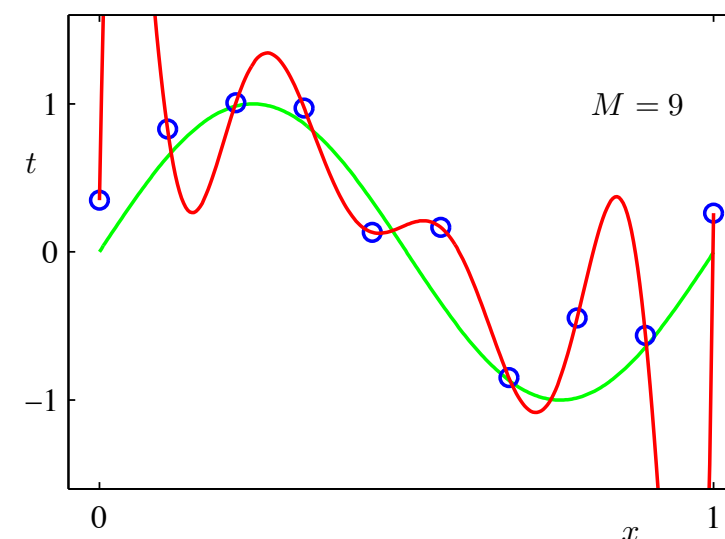
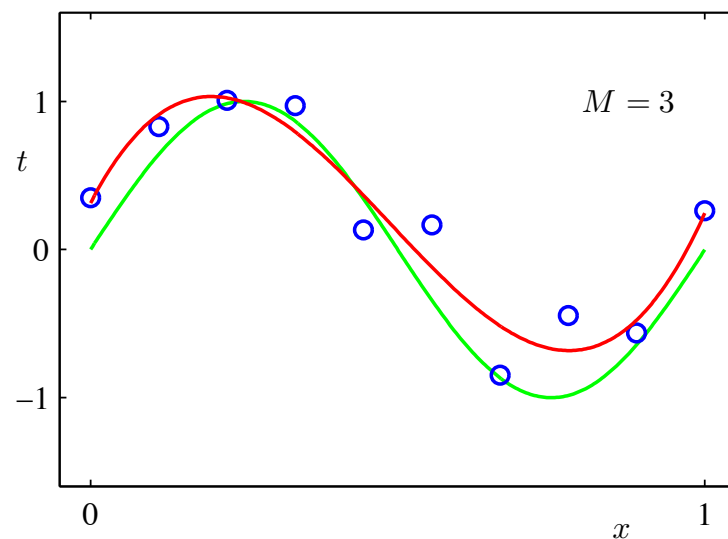
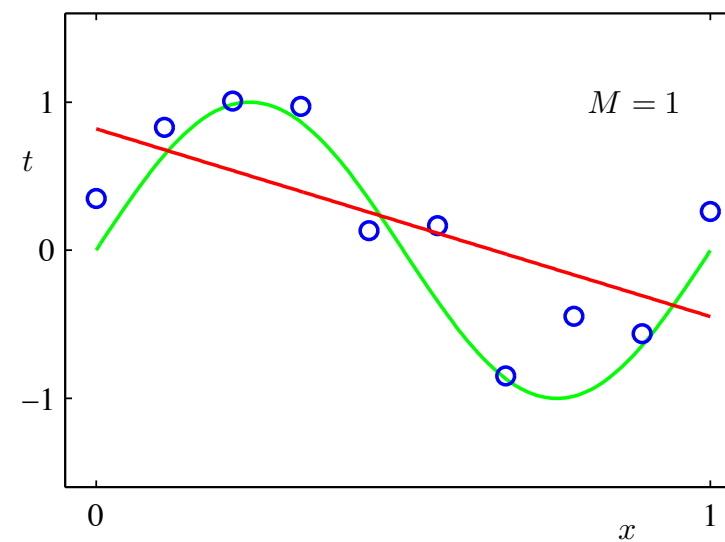
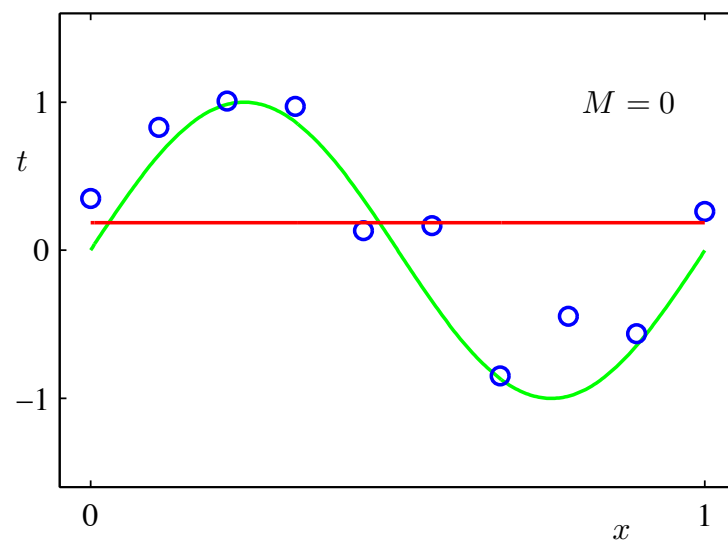
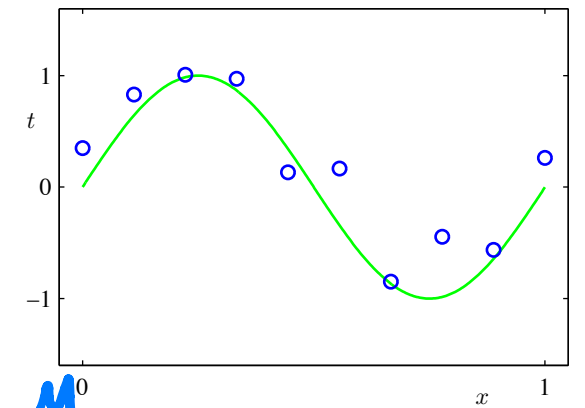
noise : $\varepsilon \sim \mathcal{N}(0, 1)$



$y(x)$

T: Class of tasks

Regression $f(x) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M$



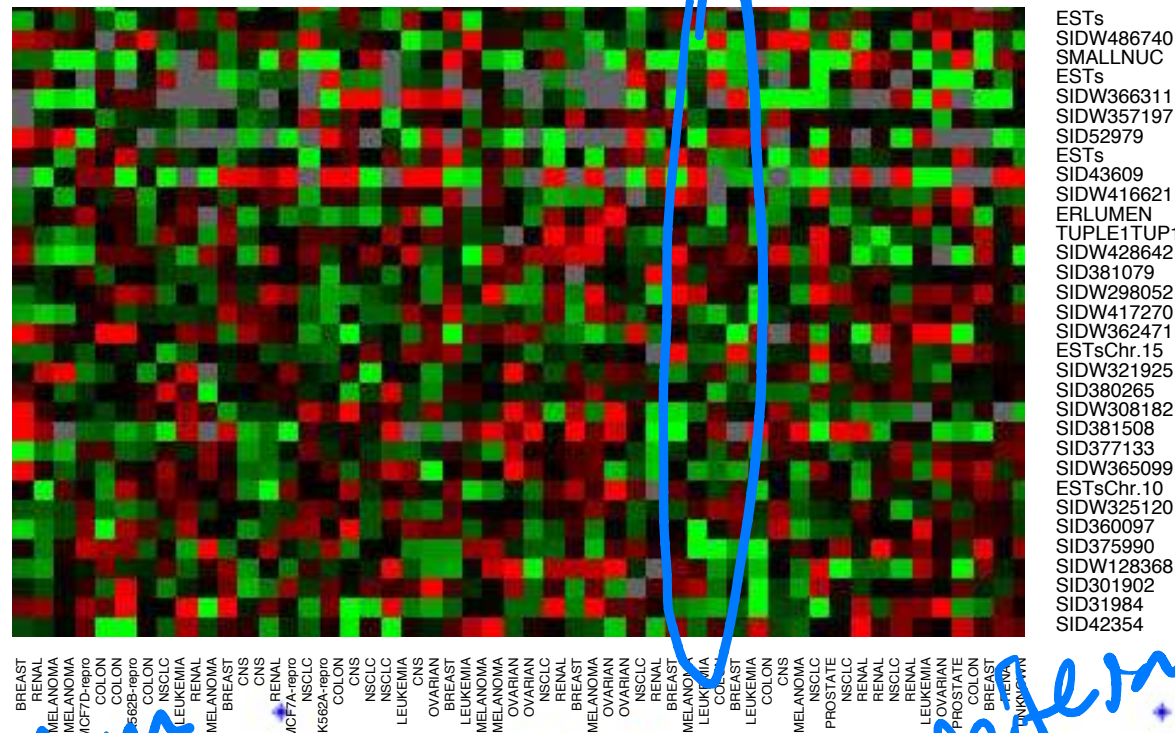
Polynomials of order M (red) fit to data constructed as $t = \sin(2\pi x) + \varepsilon$ (green)

T: Class of tasks

Clustering

N-dim
vector

20 point



(most similar)

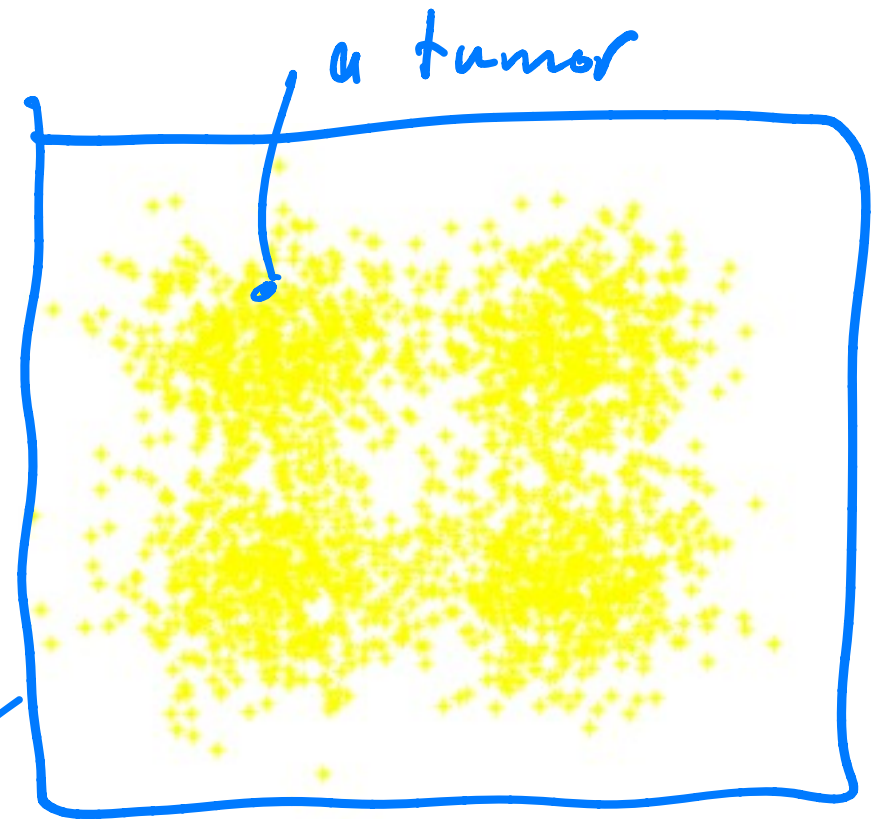
look ↓

reference

look for closest cluster

LEUKEMIA
OVARIAN
PROSTATE
COLON
BREAST
PANCREAS
LUNGADEN

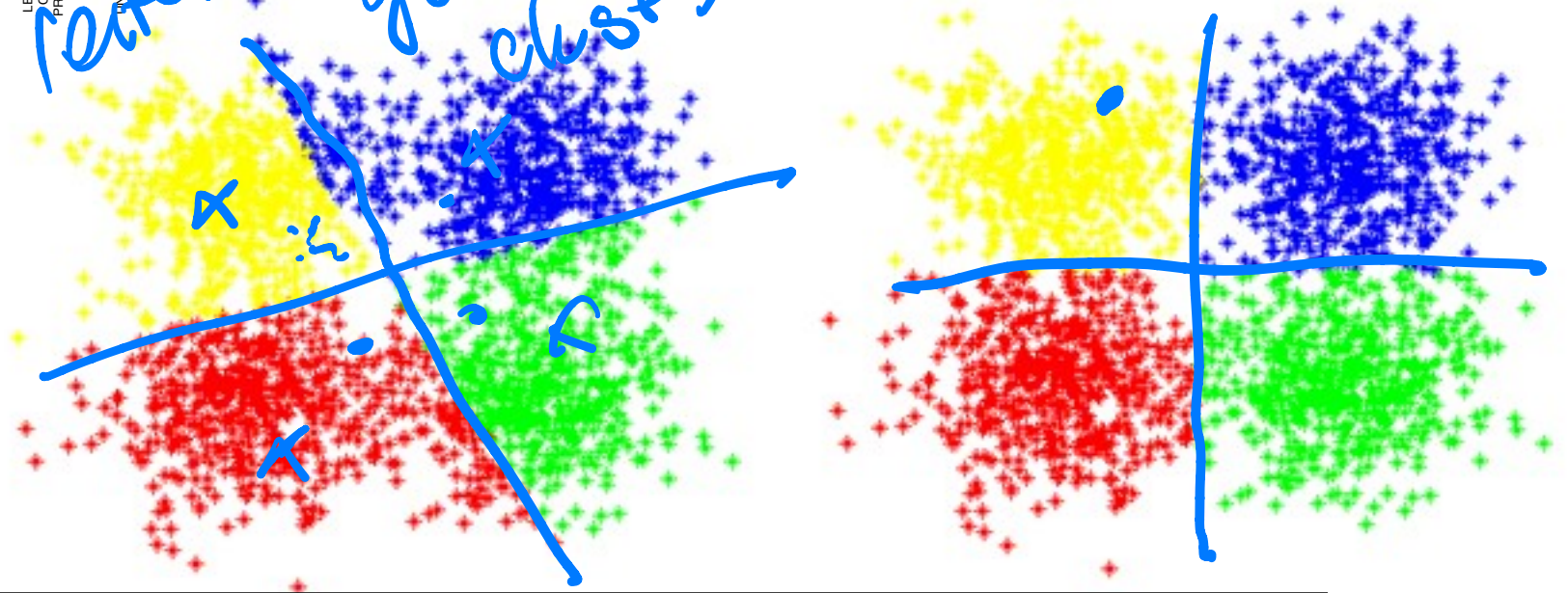
EST1539r.10
SIDW325120
SID360097
SID375990
SIDW128368
SID301902
SID31984
SID42354



random
assign
each
 x_i to
one of
4 clusters






Exp



Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

P: Performance measure

Classification

 = 2
 = 8
 = 0

All caps Trigger phrase Price
 Time is running out Save 50% on all the best moments! - 50% Off Photo Purchase, \$3.99 T
 you're so close to FREE snacks! - we love you to try our delicious snacks | graze claim your
 Last Chance: Start 2016 with 50% off .com - Get more from your 2016 with .com -
 Additional Incentive - Great news Elise, from now through the end of the month is offeri
 PROOF: Diabetes Reversed 100% Naturally - To receive this email in your inbox and activate t
 Exclamation point Attachment

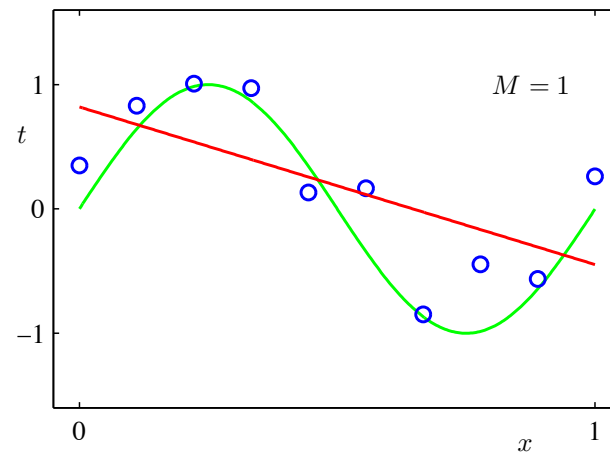
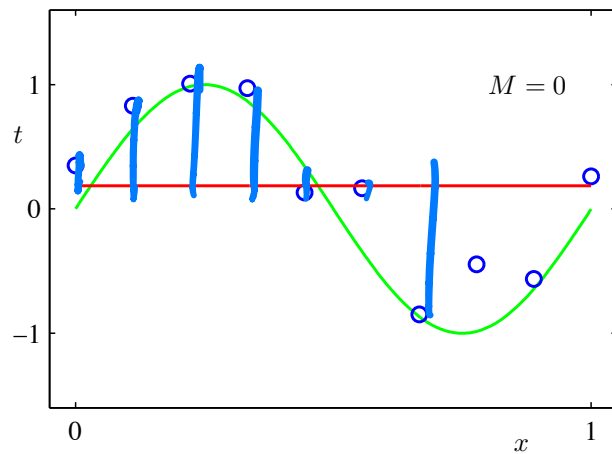
$$\text{accuracy}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} \mathbb{1}[y_i = \hat{y}_i]$$

indicator funct. $\approx \begin{cases} 1 & \text{if } y_i = \hat{y}_i \\ 0 & \text{otherwise} \end{cases}$

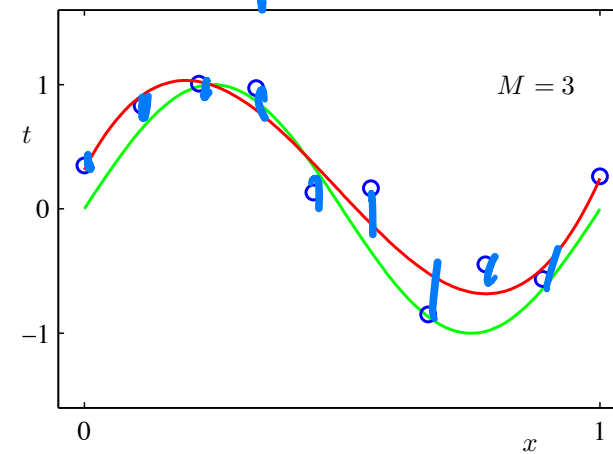
P: Performance measure

Regression

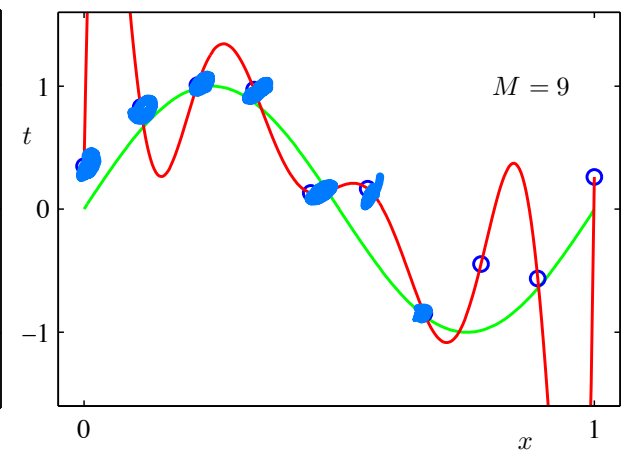
large MSE



small MSE



MSE = 0



mean squared error

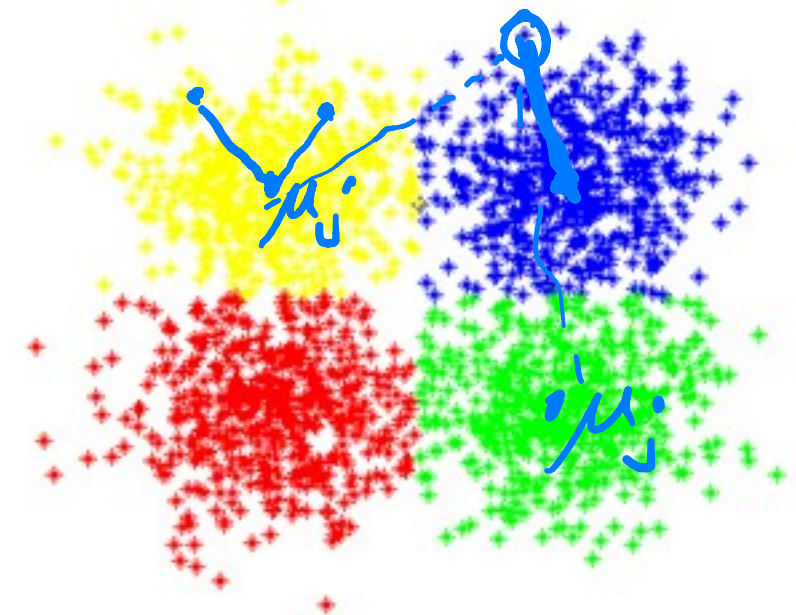
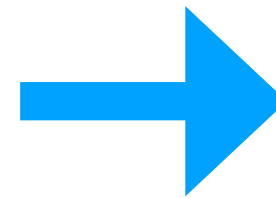
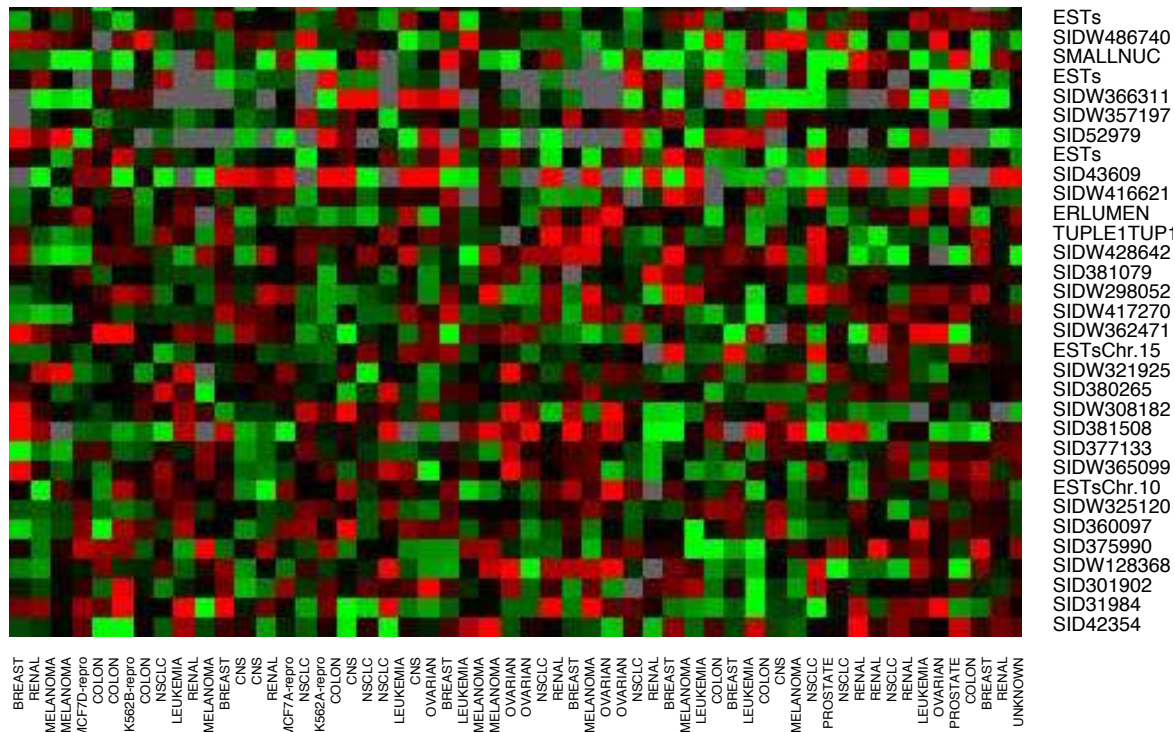
$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2$$

$\hat{y}_i = f_{\underline{w}}(x_i)$

Polynomials of order M (red) fit to data constructed as $t = \sin(2\pi x) + \varepsilon$ (green)

P: Performance measure

Clustering

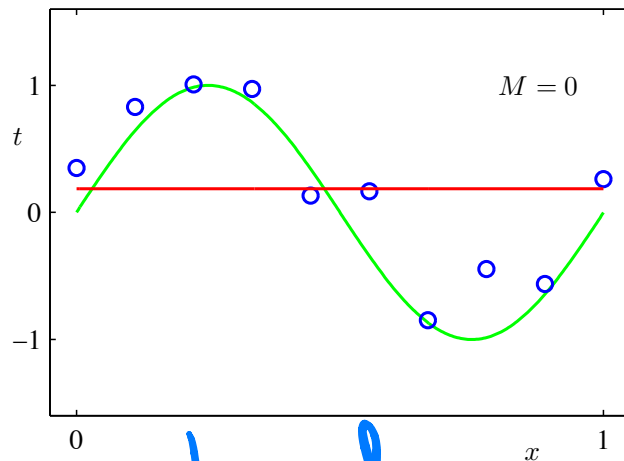


within cluster sum of squares =
$$\sum_{i=1}^{n_{\text{samples}}} \min_{\mu_j \in C} \|\mu_j - x_i\|^2$$

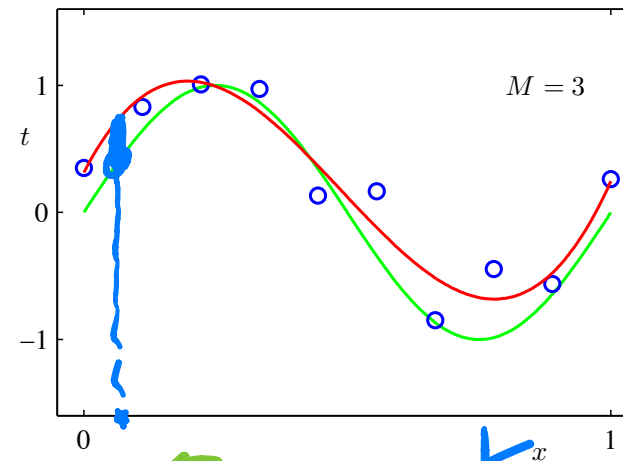
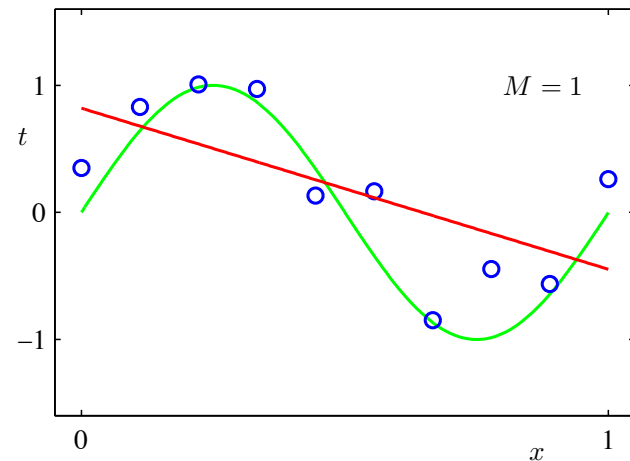
Expression matrix of genes (rows) for 64 human tumor samples (columns). [source: ESL 1.3]

P: Performance measure

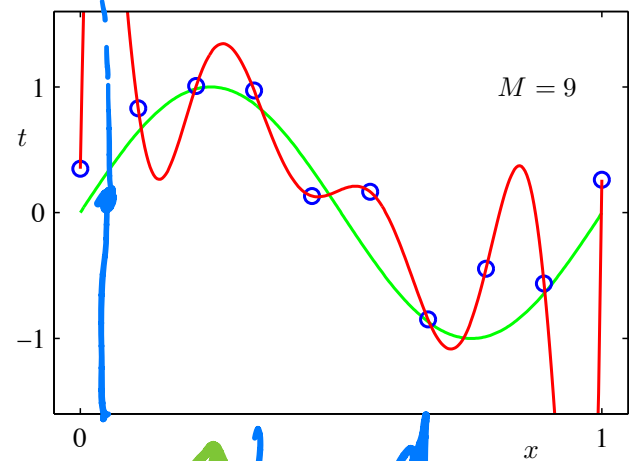
$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2$$



bad



great



overfitting

bad

Best performance on training set :

Best performance on new datapoints :

Q: On which datapoints should performance be measured?

Generalisation:

performance should be
measured on new data
(test data)

What is machine learning?

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E .”

- Tom M. Mitchell